RS LIBRARY - OGDEN CENTER
MONTHLY ALERT
Edition July 21 2002
Item No. RMF
File

Final Report

For

PROVIDED BY
THE RMRS LIBRARY
MATERIAL MAY BE PROTECTED BY
COPYRIGHT LAW, TITLE 17 US CODE

"Extrapolating modeling results from SIMPPLLE and MAGIS to Like Areas"
Phase I
(RMRS-00-JV-11222052-527)

Prepared
By

Dr. Hans Zuuring
School of Forestry
The University of Montana
Missoula, MT 59812
(406-243-6456)

February 18, 2002

21. Zuuring, Hans. *Extrapolating modeling results from SIMPPLLE and MAGIS to like areas : Phase I.* 2002: 13 p. [Unpublished final report ; RMRS-00-JV-11222052-527]

STUDY OVERVIEW

In the past, long-range forest plans have, for the most part, been non-spatial, that is, without proposed practices or treatments being applied to specific locations in specific timeframes. This has resulted in an inability to fully implement plans because treatments and associated resource effects are often unacceptable when planned treatments are proposed for implementation in real time and space. Conducting spatial analyses on a small number of "representative areas" selected in the forest planning process would help determine whether the desired future conditions targeted in a non-spatial strategic plan can be achieved, and if so over what time frame. It would also improve quantification of: 1) extent of natural disturbance factors with and without plan treatments, 2) associated effects on various resource values, and 3) expected costs. If the results from these representative areas can be extrapolated to other "like areas," then even more information could be efficiently developed to assess the feasibility of implementing a non-spatial strategic plan and to provide feedback for modifying plans.

OBJECTIVES

The overall objective of this study is to test whether spatially explicit model results from a "representative area" can be extrapolated to "like areas" to predict the outcomes of spatial analyses on those areas. The specific objectives associated with Phase I were:

1. Identify attributes and develop methods that permit a forested landscape to be delineated into areas that possess a common set of characteristics for the purpose of extrapolating MAGIS and SIMPPLLE modeling results from representative areas to "like" areas.

2. Build and run MAGIS and SIMPPLLE models for a representative area.

Objective 1 Section:

METHODS

Data Collection:

1) Geographic

In order to characterize landscapes and form clusters of similar landscape units, a 5th – order **h**ydrologic **u**nit **c**ode (HUC - a naturally determined hierarchy of watersheds starting with the largest and working back to the smallest) watershed was chosen as the sampling unit. Two ecological vegetation zones (Z8 and Z17) located east of the continental divide in Montana were selected as suitable data collection sites. Forty-two 5th –order HUCs were identified, covering an area of approximately 3.5 million acres in total. On the average, each HUC was about 80,000 acres in size.

Geographical data and associated attributes were obtained from six different data sources listed in the table below. These data were a mixture of resolutions and types. In addition we became aware of the mixed ownership across this landscape that would impact on the historical use of that landscape.

Table 1. Description of Geographical datasets in terms of source, type and resolution.

Map Layer	Source	Data Type	Resolution
Watersheds (HUCs)	NRCS	Polygon	1:24,000
Topography	USGS NED	Raster	30 m^2
Vegetation	SILC 1	Polygon	5-acre MMU
Hydrography	TIGER	Line	1:100,000
Transportation	TIGER	Line	1:100,000
Climate	PRISM	Polygon	2-km^2, 2-inch depth
Fire	Fireptsr1	Point	50 – 300 ft horiz.

Watershed boundaries were based on the ten-digit hydrologic unit codes (HUCs) developed by the USDA Natural Resource Conservation Service as described at (http://www.ftw.nrcs.usda.gov/huc_data.html).

The National Elevation Data set (NED) produced by the USDI Geological Survey (http://edcnts12.cr.usgs.gov/ned/) yielded the data to produce a digital elevation model (DEM) for the area of interest. The elevation data were later used to construct numerous terrain attributes.

Vegetation attributes were derived from the Satellite Image Landcover Classification (SILC 1) dataset produced by the Wildlife Spatial Analysis Laboratory located on the University of Montana campus (http://www.wru.umt.edu). The classification of remotely sensed data for the Northern Region of the USDA Forest Service was a large on-going process and continues to be refined, i.e. SILC 2 and SILC 3. This data consisted of 18 overlapping ESRI ArcInfo Grids derived from 18 LANDSAT Thematic Mapper scenes imaged between July 6, 1991 and September 25, 1993.

In order to maintain a consistent scale across the entire study area, both stream and road data files were obtained from the Census Bureau's 1995 TIGER (Topologically Integrated Geographic Encoding and Referencing) database of geographic information (http://mapserver4.esri.com/ado1/tiger/). The database contained information about streams and roads such as their location in latitude and longitude, name, feature type, street address ranges, geographic relationship to other features, and other attributes.

Climate was characterized as an area-weighted estimate of mean annual precipitation depth (2 inches). A polygon coverage of average annual precipitation for the climatological period 1961 – 90 was used for these calculations. Polygons were created

from Parameter-elevation Regressions on Independent Slopes Model (PRISM) derived raster data. PRISM is an analytical model that uses point data and a digital elevation model (DEM) to generate gridded estimates of annual, monthly and event-based climatic parameters (http://www.ocs.orst.edu/prism/prism_new.html).

Fire in the Rocky Mountains is a process that, in part, dictates other processes on the landscape such as insect and disease outbreaks. A Fire Starts coverage produced for the Northern Region of the USDA Forest Service by the Flathead National Forest in Kalispell, MT was used to characterize a landscape's propensity to burn. This point coverage contains the location of all R1 USFS fire records (USFS lands only) taken from fire reporting databases for the period 1960 – 2000 (Resource Information Management Coordinator, Don Krogstad, 1935 3rd Ave. East, Kalispell, MT 59901. Phone: (406) – 758-5223 E-mail: dkrogstad@fs.fed.us).

2) Attributes

Variables based on *geophysical processes*, *vegetation*, and *past management activities* were judged to be important for characterizing these 5th -order HUCs. A total of **101** variables were assembled for each of the 42 HUCs. These variables were either imported or derived. They were organized into nine groups namely topography, networks, watershed descriptors, climate, fire, landuse, vegetation, landscape metrics, and human impacts (see Table 2).

Table 2. Organization of original variables.

Category	Description	# of vars	Source
Topography	Elevation, slope aspect	19	Derived
Networks	Road/stream length	2	Derived
Watershed	Area, perimeter, road & stream density, circularity, wetness, streampower, order	11	Derived
Climate	Precipitation	3	Derived
Fire	# of starts, area burned	5	Imported
Landuse	Forest, non-forest, water, non-veg, agriculture	5	Imported
Vegetation	Cover by dominant tree species and habitat type	29	Imported
Landscape	Shannon's & Simpson's diversity and evenness indices, patch richness by species, size class & density.	21	Derived
Human	Well # and density and population	6	Imported

Data Analyses:

Through exploratory data analyses (EDA – graphical displays and correlation coefficients) statistically important variables were identified. Using repeated trials of either a hierarchical or K-means cluster analysis (multivariate statistical technique) based on different combinations of independent variables, a common appropriate clustering method, and interval-scale clustering metric, a number of homogeneous clusters of 5^{th} – order HUCs were identified. The ideal number of clusters would seem to be four with an approximate equal number of HUCs in each since there are only 42 watersheds in total. In association with a hierarchical cluster analysis a dendrogram and icicle diagram were produced for each trial to help decide which models produced meaningful results. Since no statistics are generated with this clustering approach the K-means cluster analysis was also conducted because an ANOVA table is generated.

Further, a subsequent discriminant analysis was run for each trial to test the reliability and validity of the associated cluster analysis. A cross-validation (leave-one-out classification) table and territorial map (display of discriminant scores of function 1 and 2 with group centroids) were produced. In addition, the predicted group membership, discriminant scores, and probabilities of group membership were saved in an output file for later processing.

From the territorial map the classified HUC closest in Euclidian distance to the centroid of each group was selected as the *representative* area and two other HUCs were selected at random from each group to be the 'like' areas.

SPSS v10.0 was used to perform all statistical analyses and Microsoft EXCEL was used to compute Euclidian distances between group centroids and HUC discriminant scores from the first two functions, graph a better territorial map, and assist in identifying the representative and like areas from each cluster.

RESULTS

EDA as well as repeated trials of hierarchical cluster analysis led to the identification of nine statistically important independent variables namely:

- Median elevation and elevation range (**DEM_RANGE** & **DEM_MEDIAN**)
- Average slope (**SLOPE_MEAN**)
- Majority Aspect (**ASPECT_MED**)
- Percent of HUC area forested (**FOREST**)
- Percent lodgepole pine cover (**LP**)
- Number of fire starts (**STARTCOUNT**)
- Average precipitation (**AVE_IN**)
- Shannon Diversity Index (**SC_SHDI**)

The default *Between-groups linkage* method was chosen out of a set of six possible clustering techniques using the default *Squared Euclidian Distance* as the chosen metric (out of a possible set of seven). When the dendrograms were examined it was very difficult to determine the validity of the resultant clustering and deciding what was an appropriate number of clusters. Given that no statistics were reported K-means cluster analyses were conducted instead because an ANOVA table could be generated. This alternative procedure did require the specification of the number of clusters (4 was chosen) and no choices were provided regarding the clustering method or clustering metric. The same defaults as specified in the hierarchical cluster analysis were used. Although the F-values in the ANOVA table are opportunistic (the procedure tries to form groups that do differ), the relative size of these statistics provides information about each variable's contribution to the separation of the groups.

The final acceptable grouping based on nine independent variables is given in Table 3 and the ANOVA results are presented in Table 4.

Table 3. Final cluster membership.

Number of Cases in each Cluster

Cluster	1	15.000
	2	4.000
	3	16.000
	4	7.000
Valid		42.000
Missing		.000

Table 4. F-values for nine independent variables based on 4 clusters.

ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
DEM_RANGE	527280.137	3	13615.107	38	38.728	.000
DEM_MEDIAN	484568.991	3	20977.621	38	23.099	.000
SLOPE_MEAN	62.249	3	12.611	38	4.936	.005
ASPECT_MED	4094.466	3	902.448	38	4.537	.008
AVE_IN	163.104	3	7.057	38	23.111	.000
STARTCOUNT	2360.738	3	895.626	38	2.636	.064
FOREST	6553.492	3	310.708	38	21.092	.000
LP	4096.807	3	140.144	38	29.233	.000
SC_SHDI	.166	3	3.759E-02	38	4.404	.009

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means ar equal.

The number of HUCs present in each cluster ranged from 4 to 16 (Table 3) and the F-values indicated that all nine independent variables contributed significantly towards the separation of the four clusters (Table 4).
A summary of the discriminant functions is listed in Table 5. From these results it can be seen that the first function accounts for almost 71 percent of the total variation while the first two functions account for more than 99 percent of the total variation in cluster membership (Table 5).

Table 5. Summary of discriminant functions.

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	6.294[a]	70.8	70.8	.929
2	2.527[a]	28.4	99.2	.846
3	.072[a]	.8	100.0	.259

a. First 3 canonical discriminant functions were used in the analysis.

The standardized canonical discriminant function coefficients are shown in Table 6. Focusing only on the first two functions it can be seen that DEM_MEDIAN, SLOPE_MEAN, FOREST, LP, and SC_SHDI are the dominant discriminators because the signs on these coefficients flip and their magnitude of their difference is large (Table 6). Notice that there are only three discriminant functions to separate four groups.

Table 6. Summary of three discriminant functions.

Standardized Canonical Discriminant Function Coefficients

	Function		
	1	2	3
DEM_RANGE	1.014	.550	.426
DEM_MEDIAN	1.204	-1.044	.452
SLOPE_MEAN	-1.843	.413	-1.094
ASPECT_MED	-.328	-.213	.265
AVE_IN	-.088	.275	-.767
STARTCOUNT	.123	-.184	.038
FOREST	2.014	-.865	1.422
LP	-1.334	.658	-.630
SC_SHDI	-.730	.495	.543

The highest *positive* loadings (pooled within-group correlations) associated with discriminant function 1 were observed for **DEM_RANGE** and **AVE_IN**. For discriminant function 2 the highest loadings were all *negative* for **LP**, **FOREST**, **DEM_MEDIAN**, **ASPECT_MED**, **SLOPE_MEAN**, and **STARTCOUNT** (Table 7).

The values of the discriminant functions at the cluster centroids are given in Table 8. Examination of the values associated with the first two functions indicate good separation between group centroids (Table 8).

Table 7. Loadings between independent variables and discriminant functions.

Structure Matrix

	Function		
	1	2	3
DEM_RANGE	.616*	.515	.139
AVE_IN	.528*	-.155	-.392
LP	.473	-.597*	-.111
FOREST	.353	-.590*	.193
DEM_MEDIAN	.439	-.491*	.048
ASPECT_MED	.020	-.374*	.197
SLOPE_MEAN	.167	-.291*	.127
STARTCOUNT	.086	-.253*	.080
SC_SHDI	-.159	.249	.659*

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

Table 8. Values of canonical discriminant functions at group centroids.

Functions at Group Centroids

	Function		
Cluster Number of Case	1	2	3
1	2.137	-1.485	-4.66E-02
2	3.658	3.189	.419
3	-1.377	1.004	-.235
4	-3.522	-.935	.398

Unstandardized canonical discriminant functions evaluated at group means

There are four linear classification functions, one for each cluster, and their coefficients are shown in Table 9. Notice that the magnitude of the coefficients for DEM_MEDIAN, FOREST, LP, and SC_SHDI steadily decrease from cluster 1 to 4 (Table 9).

An improved territorial map was constructed in Microsoft EXCEL. This graph clearly shows the separation between the four clusters in two-space. Cluster boundaries have been drawn as well as group centroids (Figure 1).

The discriminant scores associated with each cluster member and their group centroid were then used to calculate Euclidian distances and the HUC member closest to the group

centroid was chosen as the *representative* area. The two *like* areas were drawn from the remaining cluster members at random. These selected HUCs are listed in Table 10.

Table 9. Classification function coefficients for four clusters.

Classification Function Coefficients

	Cluster Number of Case			
	1	2	3	4
DEM_RANGE	.136	.173	.117	9.106E-02
DEM_MEDIAN	.411	.391	.363	.361
SLOPE_MEAN	-14.397	-14.787	-12.226	-11.533
ASPECT_MED	-1.24E-02	-5.81E-02	6.620E-03	4.939E-02
AVE_IN	-1.561	-1.262	-1.132	-1.444
STARTCOUNT	-2.20E-02	-4.39E-02	-5.20E-02	-4.81E-02
FOREST	3.312	3.294	2.773	2.674
LP	-5.802	-5.739	-5.258	-5.157
SC_SHDI	-90.678	-83.176	-71.628	-66.727
(Constant)	-304.817	-331.017	-248.858	-231.754

Fisher's linear discriminant functions

Figure 1. Territorial map of first two discriminant functions showing cluster boundaries, HUC members, and group centroids

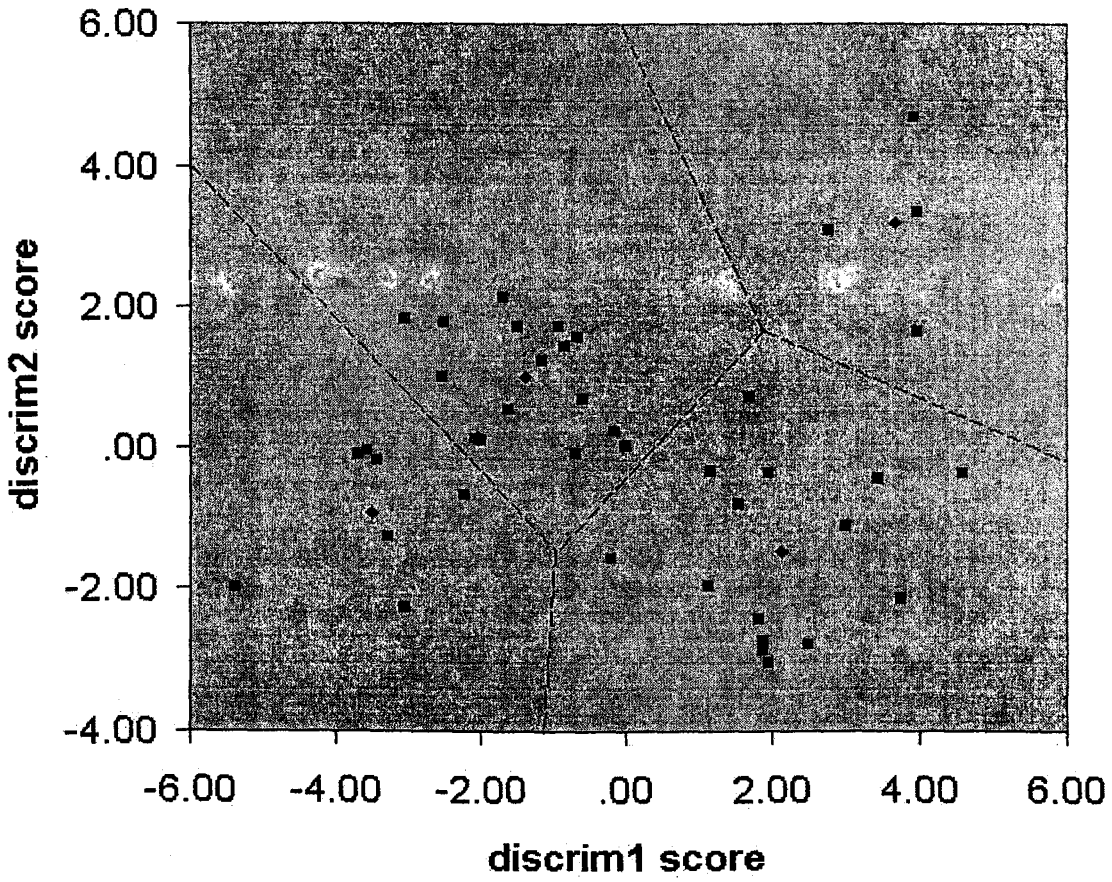

Table 10. Summary of chosen 5th-order HUCs.

FIVECODE	Cluster_id	Finalists
1003010501	1	
1004010310	1	
1003010309	1	
1004010301	1	Like1
1003010307	1	
1701020211	1	
1701020212	1	**Rep**
1701020202	1	
1701020203	1	
1701020210	1	
1701020201	1	
1701020209	1	
1701020207	1	
1701020208	1	
1701020509	1	Like2
1701020109	2	**Rep**
1701020105	2	
1701020104	2	Like1
1701020103	2	Like2
1003010313	3	
1004010201	3	
1003010503	3	
1003010502	3	Like1
1004010202	3	
1003010310	3	Like2
1004010302	3	
1004010303	3	
1701020206	3	
1003010301	3	
1004020105	3	
1004020103	3	
1701020205	3	
1701020204	3	**Rep**
1003010302	3	
1004020101	3	
1003010211	4	
1004010308	4	
1003010306	4	Like1
1701020214	4	**Rep**
1004020102	4	
1004020110	4	
1003010304	4	Like2

Finally a cross-validation table was constructed which summarized the number of HUCs correctly and incorrectly classified. Totals as well as percentages are presented. Overall, **81** percent [100 * (12 + 3 + 14 + 5) / 42] of the grouped HUCs were correctly classified. **80.0**, **75.0**, **87.5**, and **71.4** percent of the HUCs were correctly classified in clusters # 1 through 4 respectively (Table 11).

Table 11. Cross-validation of actual versus predicted group membership

		Predicted Group Membership				
	Cluster #	1	2	3	4	Total
Count	1	**12**	2	1	0	15
	2	1	**3**	0	0	4
	3	0	0	**14**	2	16
	4	0	0	2	**5**	7
Percent	1	**80.0**	13.3	6.7	0.0	100.0
	2	25.0	**75.0**	0.0	0.0	100.0
	3	0.0	0.0	**87.5**	12.5	100.0
	4	0.0	0.0	28.6	**71.4**	100.0

CONCLUSIONS

The 5th-order HUCs located in two ecological vegetation zones situated east of the continental divide in Montana were successfully grouped into four clusters characterized by nine independent variables based on *geophysical processes*, *vegetation*, and *past management activities*. The separation between clusters of HUCs was distinct. A subsequent cross validation revealed that 81 percent of the HUCs were correctly classified.

The only reservation about these analyses comes from the fact that some of the chosen independent variables may change over time such as those associated with vegetation due to climate change. It is not known the extent to which trajectories over time associated with these discriminator variables will change. However, this concern may not be important if projection periods in subsequent planning scenarios solved by MAGIS are kept to 20 years or less.

Objective 2 Section:

This objective was partly completed. Kevin Hyde, a graduate student hired for the specific purpose of conducting various MAGIS runs, was becoming familiar with the structure and operation of MAGIS. In addition, Kevin gathered and organized both geographic and tabular data sets as well as formulating the types of planning scenarios

that would shortly be solved by MAGIS. Kevin did run some test datasets through MAGIS and identified some program bugs that needed fixing by our computer programmers.
Because the representative and like areas were not identified until the end of December 2001, the prepared MAGIS planning models were not run. It is anticipated that these runs will occur in February 2002.